Statistical Inference Course Project (Part 2)

BELMOUIDI Mohamed

28 December, 2024

Course Project

Basic Inferential Data Analysis

Instructions

- Load the ToothGrowth data and perform some basic exploratory data analyses
- Provide a basic summary of the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there are other approaches worth considering.)
- State your conclusions and the assumptions needed for your conclusions.

Exploratory Data Analysis

First, we load the required packages and the dataset:

```
## Warning: ‘tbl_df()‘ was deprecated in dplyr 1.0.0.
## i Please use ‘tibble::as_tibble()‘ instead.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was generated.
```

We examine the structure and summarize the dataset:

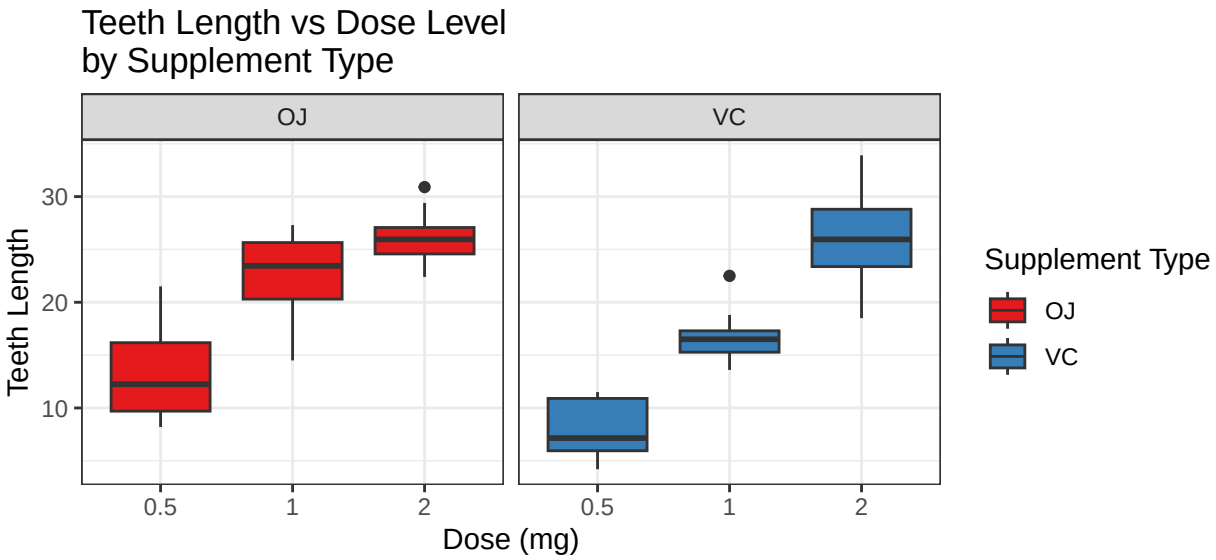
```
## tibble [60 x 3] (S3: tbl_df/tbl/data.frame)
##  $ len : num [1:60] 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

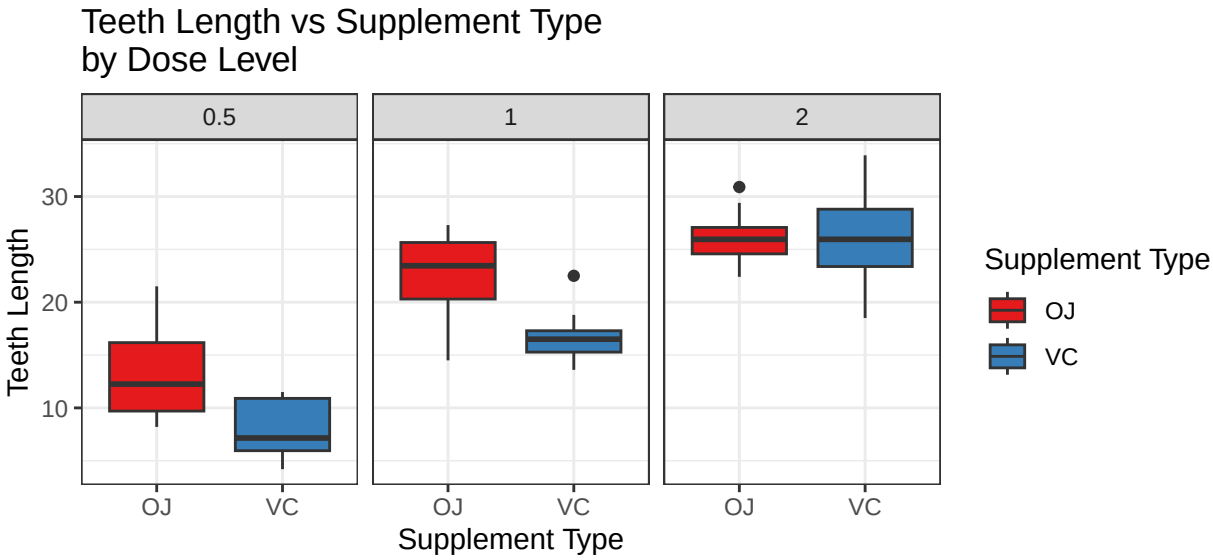
We observe the unique values of the `dose` variable:

```
## [1] 0.5 1   2
## Levels: 0.5 1 2
```

The `dose` variable contains three unique values (0.5, 1, 2). We convert it to a factor:



Plots



Hypothesis Tests

We conduct statistical tests to compare teeth growth by supplement type and dose levels.

Test by Dose Levels

-

```
##
##  Welch Two Sample t-test
##
## data:  len_a by dose_a
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means between group 0.5 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

-

```
##
##  Welch Two Sample t-test
##
## data:  len_b by dose_b
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means between group 0.5 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

-

```
##
##  Welch Two Sample t-test
##
## data:  len_c by dose_c
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

Test by Supplement Type

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Conclusions

Key Findings:

1. There is a statistically significant relationship between dose levels and teeth length (p-values < 0.05 in all dose comparisons).
2. No statistically significant difference exists between supplement types (p-value $= 0.06$).

Assumptions for Validity:

- Independence of samples: Random sampling or assignment.
- Normality of population distributions.

These assumptions are critical to ensure the validity of t-test results.